

Addressing Racial and Ethnic Disparities in Health Care: Using Federal Data to Support Local Programs to Eliminate Disparities

Thomas D. Sequist and Eric C. Schneider

To reduce racial and ethnic disparities in health care, managers, policy makers, and researchers need valid and reliable data on the race and ethnicity of individuals and populations. The federal government is one of the most important sources of such data. In this paper we review the strengths and weaknesses of federal data that pertain to racial and ethnic disparities in health care. We describe recent developments that are likely to influence how these data can be used in the future and discuss how local programs could make use of these data.

Key Words. Federal government, health disparities, racial disparities, databases

Racial and ethnic disparities in both the use of health care services and the quality of care within the United States have been well documented (Gornick et al. 1996; Institute of Medicine 2002; Virnig et al. 2002; Weech-Maldonado et al. 2003; Haas et al. 2004; Fremont et al. 2005). The recent National Health Care Disparities Report (NHDR) and other studies summarize these disparities on a national level (Schneider, Zaslavsky, and Epstein 2002; AHRQ 2005; Centers for Disease Control and Prevention 2005; Felix-Aaron et al. 2005). Collectively, these results have spurred national efforts to eliminate racial differences in care by the year 2010 (United States Department of Health and Human Services 2005b).

Much of what we know about racial and ethnic disparities has been derived from national population samples, yet many of the actions to eliminate disparities must occur in the local collaborative efforts of regional health care organizations, communities, health care institutions, and providers (Nerenz

2005). Local groups need data to guide interventions that address locally relevant disparities. Some efforts are already under way. In 2001, Aetna expanded efforts for the collection of race and ethnicity data on members, and used these data to target initiatives like the African American Preterm Labor Prevention and Breastfeeding Program (Hassett 2005).

Even as local projects are implemented, the federal government will remain a major source of data for monitoring racial and ethnic health care disparities (Lurie, Jung, and Lavizzo-Mourey 2005). In this paper we summarize the key features of selected federal health care databases. We examine their strengths and weaknesses, as well as trends that may affect their future use. Finally, we propose a model that would enable local initiatives on health care disparities to make use of federal health data.

SUMMARY OF FEDERAL SOURCES OF RACE AND ETHNICITY DATA FOR HEALTH DISPARITIES

A variety of federal programs collect data on race and ethnicity and health care (Committee on National Statistics 2004). Two federal data resources, the Social Security Administration (SSA) and the United States decennial census (United States Census Bureau 2005b), contain limited information on health care processes or outcomes, but supply the race and ethnicity data to federal databases used in analyses of health disparities. The SSA provides race and ethnicity data to other federal agencies including the Centers for Medicare and Medicaid Services (CMS). The U.S. census offers detailed population estimates and a detailed set of socioeconomic factors including race and ethnicity. These data are readily available through the U.S. census website (United States Census Bureau 2005b) and can be linked to multiple geographic levels (e.g., zip code or census block), permitting detailed geocoding analyses involving race and ethnicity (Fiscella and Franks 2001; Fremont et al. 2005; Krieger et al. 2005).

The federal agencies that offer health-related data can be divided into those that purchase or deliver health care and those that monitor health care (Table 1). Of the former, CMS accounts for the majority of databases (Centers

Address correspondence to Thomas D. Sequist, M.D. M.P.H., Brigham and Women's Hospital, Division of General Medicine, 1620 Tremont Street, Boston, MA 02120. Dr. Sequist is also with the Department of Health Care Policy, Harvard Medical School, Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA. Eric C. Schneider, M.D. M.Sc., is with the Department of Health Policy and Management, Harvard School of Public Health, Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA.

for Medicare and Medicaid Services 2005b). The latter group includes the Centers for Disease Control and Prevention, the National Institutes of Health, and the Agency for Healthcare Research and Quality (AHRQ). These federal databases vary in the racial and ethnic composition of the populations assessed, the allowance for multiple race designation, whether ethnicity is collected independent of race, how race and ethnicity are designated (individual self-report versus other mechanisms), and the smallest geographic unit available for analysis (e.g., census region, state, or smaller units).

The datasets in Table 1 are the most widely used and each is available to the public with restrictions on the release of some variables to decrease the risk that individuals can be identified. Some datasets are completely deidentified and freely accessible on the Internet (e.g., some CDC surveys), while others require formal approval from the collecting organization (e.g., CMS). Comprehensive lists of federal datasets containing information on race and ethnicity are available in the National Healthcare Disparities Report and from the Department of Health and Human Services (United States Department of Health and Human Services 2005a).

Table 2 summarizes the general types of health-related data available in the federal databases, organized into five general categories that are illustrative, and not meant to be mutually exclusive: (1) mortality, (2) preventive services, (3) management of chronic conditions, (4) quality of care measures, and (5) patient reported experiences and quality of life. To support analyses of racial and ethnic disparities, these databases must contain or be linked to other databases that include information on health status, use of health services, or health outcomes. Linking databases to one another further enhances the questions that can be addressed. For example, the individual-level HEDIS data submitted by health plans participating in the Medicare Managed Care program do not contain race and ethnicity information, yet linkage to the CMS enrollment files has allowed comparisons of the quality of care between racial and ethnic groups (Schneider et al. 2002; Virnig et al. 2002; Trivedi et al. 2005).

HISTORY AND VARIABILITY OF FEDERAL RACE AND ETHNICITY DATA

Federal definitions and methods for collecting race and ethnicity data have evolved between the time the U.S. census first began collecting race information in 1790 and the 2000 decennial census that for the first time allowed individuals to self-report either a single race or multiple racial and ethnic

Table 1: Overview of Federal Data Sources That Provide Health-Focused Race and Ethnicity Data

Sponsors	Approximate Sample Size	Race/Ethnicity Composition* (%)					Allows Multiple Race Designation of Individuals	Ethnicity Collected Independent of Race	Reporter of Race [†]	Smallest Analytic Unit [‡]	
		W	B	H	AI	As				Geo	Inst
Federal purchasers or providers											
CMS [§]											
Enrollment	41 × 10 ⁶	86	10	2	<1	1	No	No	†	Z	H
MCBS	12,000	87	10	6	1	1	No	Yes	SR	Z	P
CAHPS	19 × 10 ⁶	84	7	5	<1	3	Yes	Yes	SR	Z	P
ESRD	430,000	62	31	13	1	4	No	Yes	R	Z	C
HEDIS	4 × 10 ⁶	–	–	–	–	–		–	=	Z	P
VHA	730,000	80	13	5	1	1	No	No	R	Z	H
HRSA BPHC	11 × 10 ⁶	31	22	42	1	3	No	No	MR	Z	C
IHS	1.6 × 10 ⁶	–	–	–	100	–	No	No	R	Z	H
Federal health monitoring agencies											
CDC											
BRFSS	200,000	76	8	8	1	2	Yes	Yes	SR	S	–
NHANES	10,000	34	23	40	**	**	Yes	Yes	SR	Ce	–
NHIS	93,000	75	14	23	1	4	Yes	Yes	SR	Ce	–
NVSS	2.4 × 10 ⁶	86	12	5	<1	2	No	Yes	FD	Z	H
NIH											
SEER ^{††}	74 × 10 ⁶	67	11	19	1	7	No	No	MR	Z	H
AHRQ											
HCUP	19 × 10 ⁶	66	13	14	<1	3	No	No	MR	S	H
MEPS	37,000	78	15	25	1	4	No	Yes	SR	Ce	–

*W, white; B, black; H, Hispanic; AI, American Indian; As, Asian/Pacific Islander. Totals may exceed 100 if ethnicity was collected separately. Percentages are unweighted. (For reference purposes, the distribution of race/ethnicity in the 2000 U.S. Decennial Census was 72% W, 12% B, 13% H, 1% AI, 4% As).

^{||}Medicare HEDIS data does not contain race or ethnicity information, but can be linked to the Medicare enrollment database to obtain this information.

[†]Geo, smallest geographic unit (Ce, Census division; S, state; Z, zip code or county); Inst, smallest institutional unit (H, hospital; C, clinic; P, health plan).
[§]CMS, Centers for Medicare and Medicaid Services; Enrollment, CMS enrollment database (Arday et al. 2000); MCBS, Medicare Current Beneficiary Survey (Arday et al. 2000); CAHPS, Consumer Assessment of Health Plans Survey (Personal communication, Alan Zaslavsky, Medicare; CAHPS 2002); ESRD, Medicare End Stage Renal Disease Program (USRDS 2005); HEDIS, Health Employer Data and Information Set; VHA, Veterans Health Administration (Kressin et al. 2003); HRSA BPHC, Health Resources and Services Administration Bureau of Primary Health Care (Bureau of Primary Health Care 2004); IHS, Indian Health Service (Indian Health Service 1999); CDC, Centers for Disease Control and Prevention; BRFSS, Behavioral Risk Factor Surveillance System ([Centers for Disease Control and Prevention 2003a], personal communication, Eric Schneider, BRFSS 2001); NHANES, National Health and Nutrition Examination Survey (NHANES 2005); NHIS, National Health Interview Survey (NCHS 2005); NVSS, National Vital Statistics System (Kochanek, Murphy, and Anderson 2002); NIH, National Institutes of Health; SEER, Surveillance, Epidemiology, and End Results (NCI 2005); AHRQ, Agency for Healthcare Research and Quality; HCUP, Health Care Utilization Project (NIS 2002); MEPS, Medical Expenditure Panel Survey (MEPS 2002).

^{*}Data are obtained by linking to the Social Security Administration files.

[†]SR, self report; MR, medical record; R, registration; FD, funeral director.

^{***}Not available in public use files, but can be obtained upon special request.

^{††}The SEER program collects data from the following 13 states: AK, AZ, CA, CT, GA, HI, IA, KY, LA, NJ, NM, UT, WA.

Table 2: Types of Health-Related Data Available for Evaluation of Racial/Ethnic Disparities

<i>Data Source[†]</i>	<i>Mortality</i>	<i>Preventive Services</i>	<i>Management of Chronic Conditions</i>	<i>Quality of Care Measures</i>	<i>Patient Reported Experiences/Quality of Life</i>
CMS [‡]					
Claims files	✓	✓	✓	✓	
MCBS		✓	✓	✓	✓
CAHPS		✓		✓	✓
ESRD	✓	✓	✓	✓	✓
HEDIS		✓	✓	✓	
VHA	✓	✓	✓	✓	✓
HRSA BPHC		✓	✓	✓	
IHS		✓	✓	✓	
CDC					
BRFSS		✓	✓	✓	✓
NHANES		✓	✓	✓	✓
NHIS		✓	✓	✓	✓
NVSS	✓		✓		
NIH					
SEER	✓			✓	
SEER-Medicare	✓	✓	✓	✓	
AHRQ					
HCUP	✓		✓		
MEPS		✓	✓		✓

*Available through linkage to Medicare administrative database.

[†]The categories are illustrative of the types and usual use of data available, and are not meant to be inclusive of all data in each dataset.

[‡]Refer to Table 1 footnote for listing of abbreviations for each dataset.

backgrounds. The collection of ethnicity data has also evolved, with the first collection of Hispanic ethnicity in 1970 (Gibson and Jung 2002).

The collection of federal race and ethnicity data was transformed in 1977 by the Office of Management and Budget (OMB) Directive 15 (Office of Management and Budget 1978). This federal directive eliminated the category of "other race," and instead required collection of race in 4 categories: white; black; Asian, Asian American, or Pacific Islander; and Northern American Indian or Alaskan Native. It also required that Hispanic ethnicity be collected as a data element separate from race, with the option to collapse race and ethnicity into one variable, with Hispanic listed as a race. A subsequent revision of OMB Directive 15 in 1997 allowed for the identification of more than one race per individual (Federal Register 1997b).

These directives affected multiple national databases, including the Medicare program and all surveys conducted by AHRQ and the CDC. In the Medicare enrollment database (EDB), race was originally stored as white, black, other, or unknown based on SSA data derived at the time of application for a new or replacement Social Security card (Lauderdale and Goldberg 1996). After expanding the number of race categories in 1994 to comply with the original OMB Directive 15, the Medicare EDB contained a substantial number of patients with “other” and “unknown” race. To remedy this, 2.2 million beneficiaries with a race of “other,” “unknown,” or with a Hispanic surname or country of birth (from SSA files) were surveyed about their race and ethnicity in 1997. Approximately 40 percent of individuals responded to this survey, improving the completeness of the Medicare race and ethnicity data (Arday et al. 2000).

Despite the mandate of OMB Directive 15, federal race and ethnicity data still vary in several ways (Table 1). First, the relative distribution of the race and ethnicity of populations differs across databases, particularly if minority populations were deliberately oversampled. For example, while the U.S. census estimates that 12 percent of the population report black race, 23 percent (unweighted) of the National Health and Nutrition Examination Survey sample reports black race because of deliberate oversampling. Second, collection of multiple race data on individuals is not yet routine except for the U.S. Census and the surveys conducted by the CDC. The majority of other data sources do not yet include such detail (Table 1). Third, the method for collecting ethnicity data is inconsistent among databases. The Medicare EDB includes Hispanic ethnicity as a race category, whereas Medicare surveys such as the Medicare Current Beneficiary Survey (MCBS) (Centers for Medicare and Medicaid Services 2005a) and the Medicare CAHPS[®] (CAHPS 2005) separate Hispanic ethnicity from race (Table 1). Fourth, the mechanism of assessing of an individual’s race and ethnicity varies, ranging from individual self-report to the assignment of race by clerical workers at the time of patient registration (Table 1). This variability in the method of assigning race may affect the interpretation of research related to race and ethnicity (Kaplan and Bennett 2003).

KEY METHODOLOGICAL ISSUES FOR LOCAL USE OF FEDERAL HEALTH CARE DATA

There are at least four methodological issues to consider before using federal data to assess racial or ethnic disparities in health care: (1) the validity of the

classification of individuals' race and ethnicity, (2) sample size limitations, (3) the smallest analyzable geographic or institutional unit, and (4) the availability of data on other cultural or socioeconomic characteristics of the individuals that may be important mediators of health disparities.

Researchers have examined the validity of assignment of race and ethnicity in the Medicare EDB using self-report within the MCBS as a gold standard (Arday et al. 2000). The specificity of classification of minorities within the enrollment database is very high, resulting in negative predictive values ranging from 96 percent among Hispanics to 100 percent among Asian and Pacific Islanders. The sensitivity of the EDB for identifying blacks is also high (95 percent), but the sensitivity is much lower for identifying Hispanics (39 percent), Asian/Pacific Islanders (58 percent), and American Indians (11 percent). The positive predictive values for the Medicare enrollment data range from 96 percent for blacks to as low as 78 percent for American Indians, suggesting the need to exercise caution when using Medicare datasets to evaluate care for nonblack minority populations. Similar studies of the Veterans Health Administration (VHA) data compared with patient self-report found 98 percent agreement for whites, 92 percent for blacks, 76 percent for Asians, 83 percent for Hispanics, and only 23 percent for Native Americans. Even after excluding the 36 percent of patients with missing race information in the VHA, the accuracy of classification of nonblack minority populations is not optimal (Kressin et al. 2003).

Misclassification of nonblack minority individuals may bias estimates of health status or mortality. Correcting or adjusting for this can reduce the bias. For example, using tribal documentation, the Indian Health Service (IHS) and the National Center for Health Statistics noted that the proportion of American Indians misidentified in the National Death Index ranged from 1 percent in Arizona to 30 percent in California. Using corrected estimates, the IHS has produced adjusted disease-specific mortality rates for American Indians in its most recent report (Indian Health Service 1999). The IHS database might also be used to improve the low sensitivity of the Medicare EDB for identifying Native Americans. The classification of Hispanic individuals can be refined by surname analysis (Morgan, Wei, and Virnig 2004).

Medicare health plans are uniquely positioned to assess the distribution of race and ethnicity for their enrolled populations by comparing estimates from the Medicare EDB to estimates based on self-reported race and ethnicity from surveys like the Medicare CAHPS. For a given health plan, if the CAHPS data produce an unbiased estimate of the prevalence of racial and ethnic groups, and these estimates are comparable to those in the Medicare

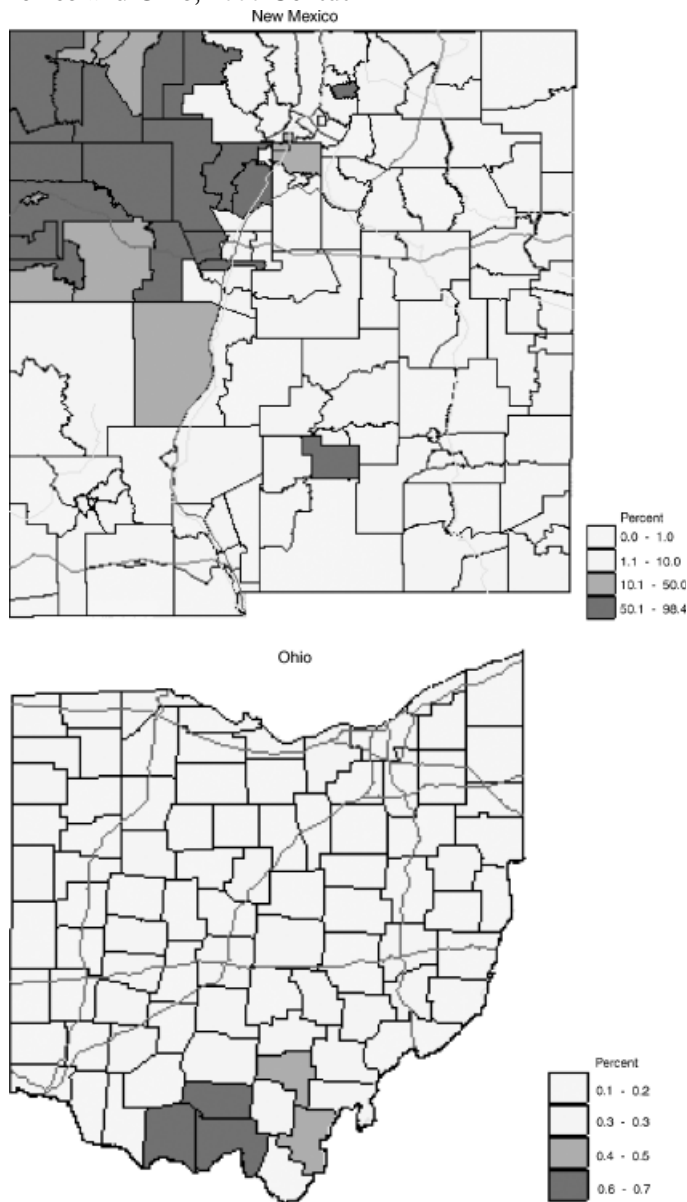
EDB, then health plan analysts could rely on the EDB to supplement race and ethnicity data on Medicare plan members. Otherwise, plans may need to request race and ethnicity information at the time of enrollment or survey members. For commercial enrollees there is no Medicare EDB equivalent.

Obtaining adequate sample sizes to reliably estimate the use of services or health status of minority populations is a second important consideration. The large numbers of individuals included in federal databases enable statistically precise estimates of many health measures. Nevertheless, the relatively small numbers of nonblack minority individuals make it difficult to measure health care delivery and outcomes with precision. Most nonblack minority groups are clustered within specific geographic regions of the country. For example, while American Indians comprise less than 1 percent of the population throughout all counties in Ohio, this proportion ranges from 1 percent to as high as 90 percent in New Mexico (United States 2005a). The utility of specific federal databases to local health care leaders will depend on the racial and ethnic composition of the local population. Fortunately, the prevalence of minority populations at the county level is readily available through maps created by the U.S. Census website (Figure 1).

Healthcare managers using federal data sources to assess local disparities will also want to know the smallest analyzable geographic or organizational unit available, such as the state, city, or individual hospital or clinic (Table 1). Most federal datasets can provide summary information at the state level and many can do so at the county or zip code level. Administrative datasets are more likely than survey datasets to achieve adequate sample sizes within smaller geographic units. Very few databases provide detailed information at the facility (hospital or clinic) level. The Dartmouth Atlas based on Medicare administrative and claims data has been used to analyze racial disparities in care based on local hospital referral regions (Baicker et al. 2004). Additional sources of facility-level data include the Medicare End Stage Renal Disease program and the AHRQ Health Care Utilization Project database. While not available routinely, the Medicare CAHPS, the MCBS, and the Medicare HEDIS program may be able to provide information on specific health plans, but whether this information is actionable will depend on the total enrollment of the health plan and the prevalence of its minority population.

Measures of socioeconomic position (SEP) such as income or education attained are useful to understand mediators of racial and ethnic disparities (Braveman et al. 2001). The primary language spoken by an individual is also increasingly recognized as an important determinant of care (Jacobs et al. 2004). There is substantial variation in the collection of both language and SEP

Figure 1: Geographic Distribution of the American Indian Population within New Mexico and Ohio, 2000 Census



The distribution of minority populations is clustered throughout the United States, as indicated by the high proportion of American Indians residing in New Mexico compared with Ohio.

variables among federal databases. Surveys conducted by the CDC can be administered in either English or Spanish, and many federal surveys include extensive SEP data, but most federal administrative and claims databases contain varying levels of information on SEP data and no information on spoken language. While it is possible to use geographic locators such as zip code within administrative or claims databases to estimate SEP through geocoding, these methods are less precise than individual level data (Krieger, Williams, and Moss 1997).

Federal databases currently lack data on other potentially important determinants of health care, such as language proficiency, health literacy, and immigrant status (Kandula, Kersey, and Lurie 2004). For example, disparities in health care within a specific racial group such as blacks may depend not only on U.S. immigrant status, but also on the precise country of origin (Lucas, Barr-Anderson, and Kington 2003; Read, Emerson, and Tarlov 2005). These subtle differences will prove important to health care organizations attempting to address disparities on a local level. For the foreseeable future, such data must be collected locally.

RELEVANT TRENDS AFFECTING THE USE OF FEDERAL RACE AND ETHNICITY DATA

In the near term, three major trends seem likely to affect the collection of race and ethnicity data and increase the complexity of their analysis: (1) the increasing diversity of the minority populations themselves, (2) the rising prevalence of non-English-speaking individuals, and (3) increasing numbers of individuals that self-identify as multiracial. Among births involving at least one black parent, the proportion with the second parent listed as white increased from 2 to 9 percent during 1968–1994 (Federal Register 1997a). The first two demographic trends tend to reduce the homogeneity of analytic categories and may increase bias due to nonresponse. In 1980, only 11 percent of the population reported speaking a language other than English at home, and this increased to 18 percent in the 2000 Census (Shin and Bruno 2003). It is likely that using survey data to assess health care disparities will underestimate existing disparities in the non-English-speaking population owing to nonresponse bias resulting from language barriers. To overcome this limitation, federal data collection efforts should translate survey instruments into additional languages as is done with the Medicare CAHPS survey. Federal agencies might also incorporate primary language spoken as a unique

administrative or claims data field to enable health assessment of the non-English-speaking populations.

Between 1982 and 1994, the proportion of patients reporting multiple races to the National Health Interview Survey (NHIS) increased from 1.2 to 1.8 percent (Federal Register 1997a), and the 2000 Census recorded 2.4 percent of respondents with multiple races (Greico and Cassidy 2001). The proportion of individuals that self-identify as more than one race also varies significantly among racial groups. For example, individuals reporting American Indian race in combination with white race accounted for 55 percent of all multiple race individuals in the NHIS analyses (Federal Register 1997a). By contrast, black race in combination with white race accounted for 11 percent of multiple race individuals in the 2000 Census (Greico and Cassidy 2001).

Allowing individuals to select up to six race designations generates 63 possible race combinations, substantially reducing the sample size within each category. Combining some categories for analysis is an option, but the optimal approach to combining them is not obvious. For example, the surveys conducted through CDC include a summary question of "Which group would you say best represents your race?" The National Center for Health Statistics has constructed statistical models that can be used to assign a single "most likely" race for multiple race individuals responding to the 2000 census. Results from these models tend to increase the relative proportions of minority populations by anywhere from 2.5 percent (blacks) to as much as 12 percent for the American Indian population (Ingram et al. 2003).

Future federal health data collection programs may enhance sampling of less numerous minority populations such as Asians and American Indians. Collection of data on nonblack minority populations might also be enhanced by refining race and ethnicity categories in special geographic locations. For example, the "Hispanic" category may not suffice in cities such as New York, where expanded definitions including Puerto Rican or Dominican might be more appropriate. CMS could play an active role in such collection efforts by defining required race/ethnicity categories to be used in each region based on the local race/ethnicity profile obtained from the U.S. Census.

A MODEL OF FEDERAL AND LOCAL COLLABORATION TO ADDRESS HEALTH DISPARITIES

Racial and ethnic disparities in health care could be addressed most effectively through collaboration between individual health care organizations and the

federal government. Any model for federal and local collaboration needs to address the division of responsibility for three data-related activities: collection, analysis, and reporting. Challenges to bidirectional sharing of standardized race and ethnicity data include time lags in the production of data, significant political and legal barriers (Kamoie and Hodge 2004), suspicions about the benevolent intent of such programs, and limited willingness to devote resources to this issue.

The federal government and states or other local agencies already collaborate to collect some forms of health care data. Three examples include collaborations between the Medicare program and the Quality Improvement Organizations (QIOs) (Jencks, Huff, and Cuerdon 2003), between the Medicare program and the National Committee for Quality Assurance and local health plans to collect standardized health plan performance measures, and between the Centers for Disease Control and Prevention and state health departments to operate the Behavioral Risk Factor Surveillance System (Centers for Disease Control and Prevention 2005a). In each instance, a local agency or organization collects data that are transmitted to the national organization for aggregation and analysis, but which are also available to the local organization for its own purposes. These or similar partnerships could be tasked with collection of standardized data on racial and ethnic health care disparities.

How would analysis and reporting on regional and local disparities be accomplished? Health care quality reporting provides two useful models. One model involves analysis and reporting of local or regional results by a national organization. For example, the National Committee for Quality Assurance (NCQA) currently generates regional and health plan summaries of HEDIS results. The other model involves reporting by regional, state, or local groups. For example, the Massachusetts Health Quality Partnership has produced statewide reports on the quality of physician groups by aggregating and analyzing the data that local health plans report to NCQA. The Medicare Quality Improvement Organizations, many of which have expertise in data collection and analysis could also play this role (Jencks et al. 2003). As CMS has included the elimination of health disparities as part of the scope of work for the QIOs, local disparity reports could be a product of the QIOs (Department of Health and Human Services 1999).

Currently, the federal government reports on racial and ethnic disparities in health care via the NHDR. Future releases of the NHDR might routinely include regional and local reporting on health disparities. Alternatively, the NHDR might generate local reports in response to requests from state

agencies, private sector organizations, and institutions about specific questions related to disparities. Much as census data guide local planners, these reports would enable health care leaders to better understand local manifestations of nationally documented health care disparities and enable institutional investment of resources to address locally relevant disparities (as opposed to disparities that may be relevant in other parts of the United States). Results of these local reports might also reveal resource limitations. For example, survey data could reveal local disparities in the availability of primary care or specialty services.

Some federal data, sampled nationally, may lack sufficient sample sizes to assess health care disparities among minority populations with particular medical conditions. Large claims databases are less vulnerable to sample size limitations than survey data. To address disparities using these surveys, the sampling schemes would have to be modified to oversample minority populations or initiate data collection in geographic areas that are not adequately represented. For example, there is already precedent within the Medicare Current Beneficiary Survey for oversampling selected populations (such as managed care enrollees) in some survey rounds. Likewise, monitoring cancer care for Native Americans has stimulated the addition of Arizona and Alaska to the SEER program. Sampling modifications could be guided by results of racial disparities analyses using larger administrative datasets.

In conclusion, racial and ethnic disparities continue to be an important national problem. Much of what we know about these disparities has been derived from federal databases. The federal government has key roles as standard setter, as data collector for federally sponsored programs, and as a data clearinghouse. The collection of data on race, ethnicity, and health must reflect the growing diversity of our population. Coordinating the efforts of states and local insurers with federal efforts to enhance and standardize race and ethnicity data collection could lead to more powerful analyses of aggregated data. With modifications, federal datasets can also be useful to local health care leaders and policy makers as they strive to reduce racial and ethnic disparities and improve care for all of the citizens of the United States.

ACKNOWLEDGMENT

This work was supported by a grant from the Robert Wood Johnson Foundation.

REFERENCES

- AHRQ (Agency for Healthcare Research and Quality). 2005. "National Healthcare Disparities Report" [accessed on July 17, 2005]. Available at <http://quality-tools.ahrq.gov/disparitiesreport/browse/browse.aspx>
- Arday, S. L., D. R. Arday, S. Monroe, and J. Zhang. 2000. "HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements." *Health Care Financing Review* 21 (4): 107–16.
- Baicker, K., A. Chandra, J. S. Skinner, and J. E. Wennberg. 2004. "Who You Are and Where You Live: How Race and Geography Affect the Treatment of Medicare Beneficiaries." *Health Affairs* (Suppl Web Exclusive:VAR33-44).
- Braveman, P., C. Cubbin, K. Marchi, S. Egerter, and G. Chavez. 2001. "Measuring Socioeconomic Status/Position in Studies of Racial/Ethnic Disparities: Maternal and Infant Health." *Public Health Report* 116 (5): 449–63.
- Bureau of Primary Health Care. 2004. "The Uniform Data System Calendar Year 2002 Data" [accessed on July 17, 2005]. Available at <http://bphc.hrsa.gov/uds/data.htm>
- CAHPS (Consumer Assessment of Health Plans Users Network). 2005. "Medicare CAHPS" [accessed on July 17, 2005]. Available at <http://www.cahps-sun.org/Products/Healthplan/MedicareCAHPS.asp>
- Centers for Disease Control and Prevention 2005. "Behavioral Risk Factor Surveillance System" [accessed on July 17, 2005]. Available at: <http://www.cdc.gov/brfss/>
- Centers for Medicare and Medicaid Services. 2005a. "Medicare Current Beneficiary Survey" [accessed on July 17, 2005]. Available at <http://www.cms.hhs.gov/MCBS/default.asp>
- . 2005b. "Centers for Medicare and Medicaid Services" [accessed on July 17, 2005]. Available at <http://www.cms.hhs.gov/>
- Committee on National Statistics (CNSTAT). 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: National Academies Press.
- Department of Health and Human Services. 1999. "Request for Proposal No. HCFA-99-001/ELH." Health Care Financing Administration.
- Federal Register. 1997a. "Recommendations from the Interagency Committee for the Review of the Racial and Ethnic Standards to the Office of Management and Budget Concerning Changes to the Standards for the Classification of Federal Data on Race and Ethnicity." Federal Register. Washington, DC: National Archives and Records Administration, pp. 36873–946.
- . 1997b. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." Federal Register, Washington, DC: National Archives and Records Administration, pp. 58782–90.
- Felix-Aaron, K., E. Moy, M. Kang, M. Patel, F. D. Chesley, and C. Clancy. 2005. "Variation in Quality of Men's Health Care by Race/Ethnicity and Social Class." *Medical Care* 43 (3 suppl): 172–81.
- Fiscella, K., and P. Franks. 2001. "Impact of Patient Socioeconomic Status on Physician Profiles: A Comparison of Census-Derived and Individual Measures." *Medical Care* 39 (1): 8–14.

- Fremont, A. M., A. Bierman, S. L. Wickstrom, C. E. Bird, M. Shah, J. J. Escarce, T. Horstman, and T. Rector. 2005. "Use of Geocoding in Managed Care Settings to Identify Quality Disparities." *Health Affairs (Millwood)* 24 (2): 516-26.
- Gibson, C., and K. Jung. 2002. *Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic Origin, 1970 to 1990, for the United States, Regions, Divisions, and States*. Washington, DC: U.S. Census Bureau.
- Gornick, M. E., P. W. Eggers, T. W. Reilly, R. M. Mentnech, L. K. Fitterman, L. E. Kucken, and B. C. Vladeck. 1996. "Effects of Race and Income on Mortality and Use of Services among Medicare Beneficiaries." *New England Journal of Medicine* 335 (11): 791-9.
- Greico, E., and R. Cassidy. 2001. *Overview of Race and Hispanic Origin: Census 2000 Brief*. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.
- Haas, J. S., K. A. Phillips, D. Sonneborn, C. E. McCulloch, L. C. Baker, C. P. Kaplan, E. J. Perez-Stable, and S. Y. Liang. 2004. "Variation in Access to Health Care for Different Racial/Ethnic Groups by the Racial/Ethnic Composition of an Individual's County of Residence." *Medical Care* 42 (7): 707-14.
- Hassett, P. 2005. "Taking on Racial and Ethnic Disparities in Health Care: The Experience at Aetna." *Health Affairs* 24 (2): 417-20.
- Indian Health Service. 1999. *Trends in Indian Health*. Rockville, MD: Department of Health and Human Services..
- Ingram, D., J. Parker, N. Schenker, J. Weed, B. Hamilton, E. Arias, and J. Madans. 2003. "United States Census 2000 Population with Bridged Race Categories." *Vital Health Statistics* 2 (135): 1-51.
- Institute of Medicine. 2002. *Unequal Treatment. Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press.
- Jacobs, E. A., D.S Shepard, J. A. Suaya, and E. L. Stone. 2004. "Overcoming Language Barriers in Health Care: Costs and Benefits of Interpreter Services." *American Journal of Public Health* 94 (5): 866-9.
- Jencks, S. F., E. D. Huff, and T. Cuerdon. 2003. "Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998-1999 to 2000-2001." *Journal of the American Medical Association* 289 (3): 305-12.
- Kamoie, B., and J. G. Hodge Jr. 2004. "HIPAA's Implications for Public Health Policy and Practice: Guidance from the CDC." *Public Health Report* 119 (2): 216-9.
- Kandula, N. R., M. Kersey, and N. Lurie. 2004. "Assuring the Health of Immigrants: What the Leading Health Indicators Tell Us." *Annual Review of Public Health* 25: 357-76.
- Kaplan, J. B., and T. Bennett. 2003. "Use of Race and Ethnicity in Biomedical Publication." *Journal of the American Medical Association* 289 (20): 2709-16.
- Kochanek, K. D., S. L. Murphy, and R. N. Anderson. 2002. "Deaths, Final Data for 2002." *National Vital Statistics Reports* 53 (5): 1-116.
- Kressin, N. R., B. H. Chang, A. Hendricks, and L. E. Kazis. 2003. "Agreement between Administrative Data and Patients' Self-Reports of Race/Ethnicity." *American Journal of Public Health* 93 (10): 1734-9.

- Krieger, N., J. T. Chen, P. D. Waterman, D. H. Rehkopf, and S. V. Subramanian. 2005. "Painting a Truer Picture of U.S. Socioeconomic and Racial/Ethnic Health Inequalities: The Public Health Disparities Geocoding Project." *American Journal of Public Health* 95 (2): 312–23.
- Krieger, N., D. R. Williams, and N. E. Moss. 1997. "Measuring Social Class in U.S. Public Health Research: Concepts, Methodologies, and Guidelines." *Annual Review of Public Health* 18: 341–78.
- Lauderdale, D. S., and J. Goldberg. 1996. "The Expanded Racial and Ethnic Codes in the Medicare Data Files: Their Completeness of Coverage and Accuracy." *American Journal of Public Health* 86 (5): 712–6.
- Lucas, J. W., D. J. Barr-Anderson, and R. S. Kington. 2003. "Health Status, Health Insurance, and Health Care Utilization Patterns of Immigrant Black Men." *American Journal of Public Health* 93 (10): 1740–7.
- Lurie, N., M. Jung, and R. Lavizzo-Mourey. 2005. "Disparities and Quality Improvement: Federal Policy Levers." *Health Affairs* 24 (2): 354–64.
- MEPS (Medical Expenditure Panel Survey). 2002. "MEPSnet Analytic Tools" [accessed on July 17, 2005]. Available at <http://www.meps.ahrq.gov/>
- Morgan, R. O., I. I. Wei, and B. A. Virnig. 2004. "Improving Identification of Hispanic Males in Medicare: Use of Surname Matching." *Medical Care* 42 (8): 810–6.
- NCHS (National Center for Health Statistics). 2005. "National Health Interview Survey 2002" [accessed on July 17, 2005]. Available at <http://www.cdc.gov/nchs/nhis.htm>
- NCI (National Cancer Institute). 2005. "Number of Persons by Race and Hispanic Ethnicity for SEER Participants 2000" [accessed on July 17, 2005]. Available at <http://seer.cancer.gov/registries/data.html>
- NHANES (National Health and Nutrition Examination Survey). 2005. "Demographics and Weights File Frequencies" [accessed on July 17, 2005]. Available at <http://www.cdc.gov/nchs/about/major/nhanes/frequency/demo.htm#RIDRETH1>
- Nerenz, D. R. 2005. "Health Care Organizations' Use of Race/Ethnicity Data to Address Quality Disparities." *Health Affairs* 24 (2): 409–16.
- NIS (Nationwide Inpatient Sample). 2002. "H-CUP Net" [accessed on July 17, 2005]. Available at <http://hcup.ahrq.gov/HcupNet.asp>
- Office of Management and Budget. 1978. "Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting." In *Statistical Policy Handbook*, pp. 37–38. Washington, DC: Office of Federal Statistical Policy and Standards, U.S. Department of Commerce.
- Read, J. G., M. O. Emerson, and A. Tarlov. 2005. "Implications of Black Immigrant Health for U.S. Racial Disparities in Health." *Journal of Immigrant Health* 7 (3): 205–12.
- Schneider, E. C., A. M. Zaslavsky, and A. M. Epstein. 2002. "Racial Disparities in the Quality of Care for Enrollees in Medicare Managed Care." *Journal of the American Medical Association* 287 (10): 1288–94.
- Shin, H., and R. Bruno. 2003. *Language Use and Speaking Ability: 2000*. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

- Trivedi, AN., A. M. Zaslavsky, E. C. Schneider, and J. Z. Ayanian. 2005. "Trends in the Quality of Care and Racial Disparities in Medicare Managed Care." *New England Journal of Medicine* 353 (7): 692–700.
- United States Census Bureau. 2005a. "American Factfinder. Thematic Maps" [accessed on July 17, 2005]. Available at http://factfinder.census.gov/servlet/ThematicMapFramesetServlet?_bm=y&_lang=en
- . 2005b. "United States Census 2000" [accessed on June 20, 2005b]. Available at <http://www.census.gov/main/www/cen2000.html>
- United States Department of Health and Human Services. 2005a. "Directory of Health and Human Services Data Resources" [accessed on July 17, 2005]. Available at <http://aspe.hhs.gov/datacncl/DataDir/index.shtml>
- . 2005b. "The Initiative to Eliminate Racial and Ethnic Disparities in Health" [accessed on July 17, 2005]. Available at <http://raceandhealth.hhs.gov/>
- USRDS (United States Renal Data System). 2005. "RenDER (Renal Data Extraction & Reference)" [accessed on July 17, 2005]. Available at http://www.usrds.org/odr/xrender_home.asp
- Virnig, B. A., N. Lurie, Z. Huang, D. Musgrave, A. M. McBean, and B. Dowd. 2002. "Racial Variation in Quality of Care among Medicare+Choice Enrollees." *Health Affairs* 21 (6): 224–30.
- Weech-Maldonado, R., L. S. Morales, M. Elliott, K. Spritzer, G. Marshall, and R. D. Hays. 2003. "Race/Ethnicity, Language, and Patients' Assessments of Care in Medicaid Managed Care." *Health Service Research* 38 (3): 789–808.